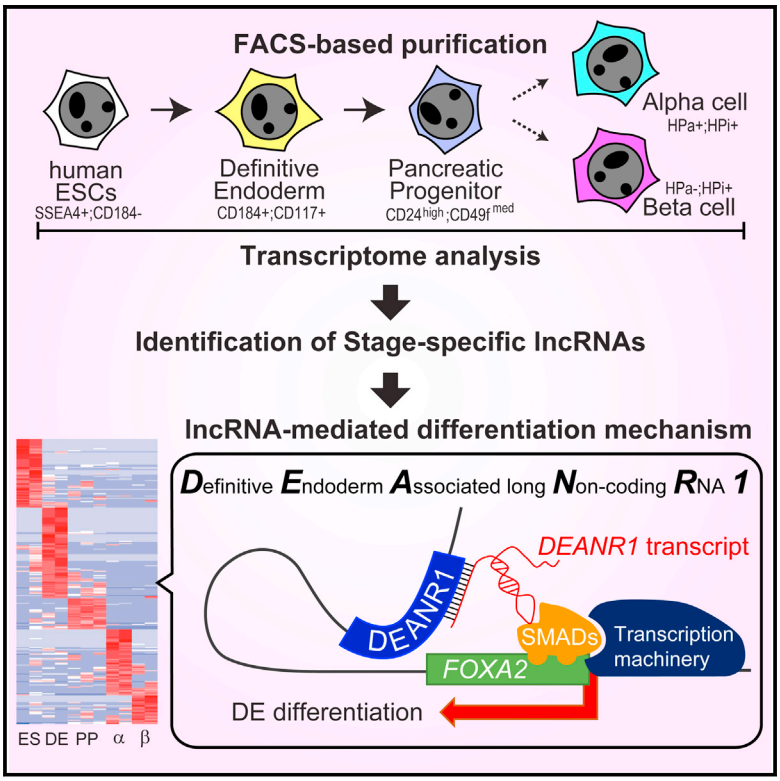
Article

Cell Reports

The lncRNA *DEANR1* Facilitates Human Endoderm Differentiation by Activating *FOXA2* Expression

Graphical Abstract



Highlights

- A transcriptome analysis of human endoderm and pancreatic lineage
- Hundreds of dynamically expressed, stage-specific lncRNAs are identified
- *DEANR1* is a functional lncRNA in human endoderm differentiation
- *DEANR1* regulates FOXA2 expression by recruiting SMAD2/3 to FOXA2 promoter

Authors

Wei Jiang, Yuting Liu, ..., Kun Zhang, Yi Zhang

Correspondence

wei.jiang@childrens.harvard.edu (W.J.), yzhang@genetics.med.harvard.edu (Y.Z.)

In Brief

Using FACS-based purification and RNA-seq, Jiang et al. comprehensively profile transcriptome dynamics during human endoderm and pancreatic lineage specification and identify hundreds of stage-specific lncRNAs. They further demonstrate that lncRNA *DEANR1* interacts with SMAD2/3 and thus functions in definitive endoderm differentiation through regulating FOXA2.

Accession Numbers

GSE44875

Jiang et al., 2015, Cell Reports *11*, 137–148
April 7, 2015
http://dx.doi.org/10.1016/j.celrep.2015.03.008

CellPress

Cell Reports
Article

The lncRNA *DEANR1* Facilitates Human Endoderm Differentiation by Activating *FOXA2* Expression

Wei Jiang,[1,2,3,*] Yuting Liu,[1,2,3] Rui Liu,[5] Kun Zhang,[5] and Yi Zhang[1,2,3,4,*]
[1]Howard Hughes Medical Institute, Harvard Medical School, WAB-149G, 200 Longwood Avenue, Boston, MA 02115, USA
[2]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Harvard Medical School, WAB-149G, 200 Longwood Avenue, Boston, MA 02115, USA
[3]Department of Genetics, Harvard Medical School, WAB-149G, 200 Longwood Avenue, Boston, MA 02115, USA
[4]Harvard Stem Cell Institute, Harvard Medical School, WAB-149G, 200 Longwood Avenue, Boston, MA 02115, USA
[5]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093-0412, USA
*Correspondence: wei.jiang@childrens.harvard.edu (W.J.), yzhang@genetics.med.harvard.edu (Y.Z.)
http://dx.doi.org/10.1016/j.celrep.2015.03.008

SUMMARY

Long non-coding RNAs (lncRNAs) regulate diverse biological processes, including cell lineage specification. Here, we report transcriptome profiling of human endoderm and pancreatic cell lineages using purified cell populations. Analysis of the data sets allows us to identify hundreds of lncRNAs that exhibit differentiation-stage-specific expression patterns. As a first step in characterizing these lncRNAs, we focus on an endoderm-specific lncRNA, definitive endoderm-associated lncRNA1 (*DEANR1*), and demonstrate that it plays an important role in human endoderm differentiation. *DEANR1* contributes to endoderm differentiation by positively regulating expression of the endoderm factor *FOXA2*. Importantly, overexpression of *FOXA2* is able to rescue endoderm differentiation defects caused by *DEANR1* depletion. Mechanistically, *DEANR1* facilitates *FOXA2* activation by facilitating SMAD2/3 recruitment to the *FOXA2* promoter. Thus, our study not only reveals a large set of differentiation-stage-specific lncRNAs but also characterizes a functional lncRNA that is important for endoderm differentiation.

INTRODUCTION

Definitive endoderm (DE) specification is one of the first critical steps in early development that generate gastrointestinal tracts and all of the associated organs/tissues, including the pancreas and liver (Zorn and Wells, 2009). Previous studies have shown that human embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) could be induced to differentiate into DE and further into pancreatic progenitors (PPs), as well as islet-like structures exhibiting a partial insulin-secretion function (D'Amour et al., 2006; Jiang et al., 2007; Nostro et al., 2011; Zhang et al., 2009). The derived PPs can be further matured to form well-structured islets that are capable of responding to glucose following transplantation in mouse models (Kroon et al., 2008). However, the in vivo maturation approach was not applicable to clinical studies or suitable for generating large quantities of relevant cells for drug screening until very recently, when two groups reported their success in making functional beta cells in vitro (Pagliuca et al., 2014; Rezania et al., 2014). Additionally, differences between human and mouse endoderm and beta cell development, such as cell-cell interaction and islet structure, have prevented direct translation of the knowledge gained from mouse studies to human applications (McKnight et al., 2010). For instance, a microarray-based transcriptome analysis of human exocrine and endocrine pancreatic cell types not only revealed enrichment of genes important for paracrine signaling but also showed an unexpected abundance of some regulatory proteins, including MAFB in beta cells, distinct from those reported in mouse (Dorrell et al., 2011). Since the study of human development is hampered by limited access to human tissues, particularly of the embryonic stage, understanding cell-lineage-specific differentiation of human ESCs provides an alternative. Indeed, a recent study utilized such a system to investigate human cardiac lineage specification (Paige et al., 2012). Therefore, it is possible that a comprehensive transcriptome characterization of the different developmental stages of endoderm and pancreatic cell lineages would generate a rich resource for understanding the regulatory networks of cell lineage specification and provide additional insights into the mechanism of beta cell differentiation.

The FANTOM and ENCODE projects have revealed that a large proportion of the mammalian genome is transcribed (Carninci et al., 2005; Djebali et al., 2012; Katayama et al., 2005). Thousands of these transcripts are classified as long non-coding RNAs (lncRNAs), with size larger than 200 nucleotides, but very few have been functionally characterized. Biochemically, lncRNAs have been shown to mediate local gene expression in *cis*, affect multiple gene transcription in *trans*, and serve as a scaffold for chromatin structure maintenance (Rinn and Chang, 2012; Ulitsky and Bartel, 2013). Functionally, lncRNAs have been shown to regulate diverse biological processes, including X chromosome inactivation, genomic imprinting, cell-cycle regulation, and regulation of stem cell pluripotency (Wang and

Chang, 2011). Moreover, lncRNAs are less conserved compared with protein coding genes of different species and exhibit tissue- and cell-type-specific expression (Cabili et al., 2011; Derrien et al., 2012). Very recently, *Fendrr*, an lncRNA that is specifically expressed in lateral mesoderm, was shown to be important for mouse heart and body wall development (Grote et al., 2013; Sauvageau et al., 2013); *Braveheart*, a mouse- and heart-specific lncRNA, was shown to be required for nascent mesoderm to cardiac fate specification (Klattenhoff et al., 2013); *TINCR*, a terminal differentiation-induced lncRNA in human epidermis, was reported to control epidermal differentiation (Kretz et al., 2013); and *lnc-DC*, an lncRNA that is exclusively expressed in human dendritic cells, was shown to regulate dendritic cell differentiation (Wang et al., 2014). Collectively, these reports support the significance of lncRNAs in cell lineage specification and embryonic development.

Endoderm specification is controlled by a number of signaling pathways and transcriptional factors (Zorn and Wells, 2009), but whether lncRNAs play any role in this process is largely unknown. Interestingly, a recent study reported that during ESC-to-endoderm differentiation, more than 60% of expressed lncRNAs exhibit coordinated expression changes with their associated protein-coding genes (Sigova et al., 2013). Given that lncRNAs are capable of contributing to the regulation of gene expression, they may have a role in endoderm and pancreatic cell lineage specification, which has yet to be defined. Here, we profiled the transcriptome and revealed the dynamic expression pattern of many lncRNAs during endoderm and pancreatic cell lineage specification. We further characterized the function and mechanism of action for an important endoderm-specific lncRNA that we named definitive endoderm-associated lncRNA1 (*DEANR1*).

RESULTS

Preparation of Purified Cell Samples along the Endoderm and Pancreatic Endocrine Cell Lineage

We previously reported a protocol by which PPs can be generated in vitro by differentiation of human ESCs through the DE stage (Jiang et al., 2011; Zhang et al., 2009). Based on available cell-surface markers of the desired cell types, we prepared purified cell samples along the beta cell lineage, including undifferentiated human ESCs, DE, PPs, and primary adult islet-derived alpha and beta cells. First, we utilized a well-known human ESC surface marker (SSEA-4) combined with a lack-of-lineage marker (CD184) to purify undifferentiated human ESCs (Figure 1A). Since CD184 and CD117 can serve as surface markers for DE (Cheng et al., 2012; Nostro et al., 2011), we sorted CD184/CD117 double-positive cells (as putative DE cells) after subjecting human ESCs to DE differentiation (Jiang et al., 2013a, 2013b). We also combined CD24 (Jiang et al., 2011) with CD49f markers (Figure 1B) and purified the $CD24^{high}/CD49f^{med}$ population as the putative PP cells (Figure 1A). Fluorescence-activated cell sorting (FACS; Figure 1B) and qRT-PCR (Figure 1C) analysis confirmed the cell identities as expected.

Since no cell-surface marker for human ESC-derived beta-like cells has been reported, we chose to purify beta cells from primary human islets based on the antibodies developed by Dorrell et al. (2008, 2011). From multiple primary human islet preparations, we purified beta cells (HPi2+/HPa1−) and alpha cells (HPi2+/HPa1+) (Figure 1D). qRT-PCR analysis of cell-type-specific factors demonstrated that the cell purification was successful (Figure 1E; *Insulin* and *MAFA* mark beta cells, and *Glucagon* and *ARX* mark alpha cells).

Expression Dynamics of Protein-Coding Genes during Endoderm and Pancreatic Cell Lineage Specification

To characterize the transcriptional changes that occur during pancreatic beta cell differentiation, we performed RNA sequencing (RNA-seq) using FACS-sorted cells from 11 samples (ES, n = 2; DE, n = 2; PP, n = 3; alpha, n = 2; beta, n = 2) and a modified SMARTseq procedure for Illumina library construction (Yamaguchi et al., 2012) (detailed in Experimental Procedures). For each sample, we generated 24–45 million raw reads, >80% of which mapped uniquely (Figure S1A). Approximately 15,416 human genes were detected under the criteria of FPKM > 1 in replicates. Cluster analysis indicated that the data are highly reproducible (Figure 2A). Inspection of well-known stage-specific markers, such as SOX2, SOX17, SOX9, PDX1, insulin, and glucagon, confirmed their expected stage-specific expression (Figure S1B). The stage-specific expressed genes were enriched for Gene Ontology terms that are consistent with the function of the cells (Figure 2B). Collectively, these results are consistent with human endoderm and pancreatic cell lineage specification and development, and confirm the reliability of the data sets.

To examine the transcriptome changes that occur between developmentally related stages, we performed a pairwise comparison. Using stringent criteria (log2 [fold change] > 1.5, p value < 0.05 and higher FPKM > 5), we identified hundreds of differentially expressed genes in each comparison. KEGG signal pathway enrichment analysis (Huang et al., 2009) revealed that cell-cycle and related pathways, including metabolism and cancer pathways, were enriched in most differentiation stages (Figure 2C), which is consistent with the observation that the proliferation capacity becomes restricted during cell-fate specification. As expected, several pathways that were previously demonstrated to be involved in pancreatic cell differentiation were enriched in the pairwise analysis. For example, genes of the TGFbeta signaling pathway were enriched in both the human ES-to-DE and DE-to-PP transitions. Interestingly, MAPK and focal adhesion pathways were exclusively enriched in the PP-to-alpha transition, while the ErbB signaling pathway was exclusively enriched in the PP-to-beta transition (Figure 2C). These results suggest that these signaling pathways might coordinate PP-to-alpha/beta cell differentiation. Although previous studies of mice and other vertebrates have reported that multiple signaling pathways are involved in endoderm and pancreatic specification (Zorn and Wells, 2009), these pairwise comparisons may provide valuable information for optimizing in vitro human pancreatic differentiation procedures.

lncRNA Expression during Endoderm and Pancreatic Lineage Specification Is Highly Stage Specific

To identify lncRNAs expressed during human endoderm and pancreatic lineage specification, we mapped RNA-seq data against the GENCODE v14 lncRNA annotation (Derrien et al.,

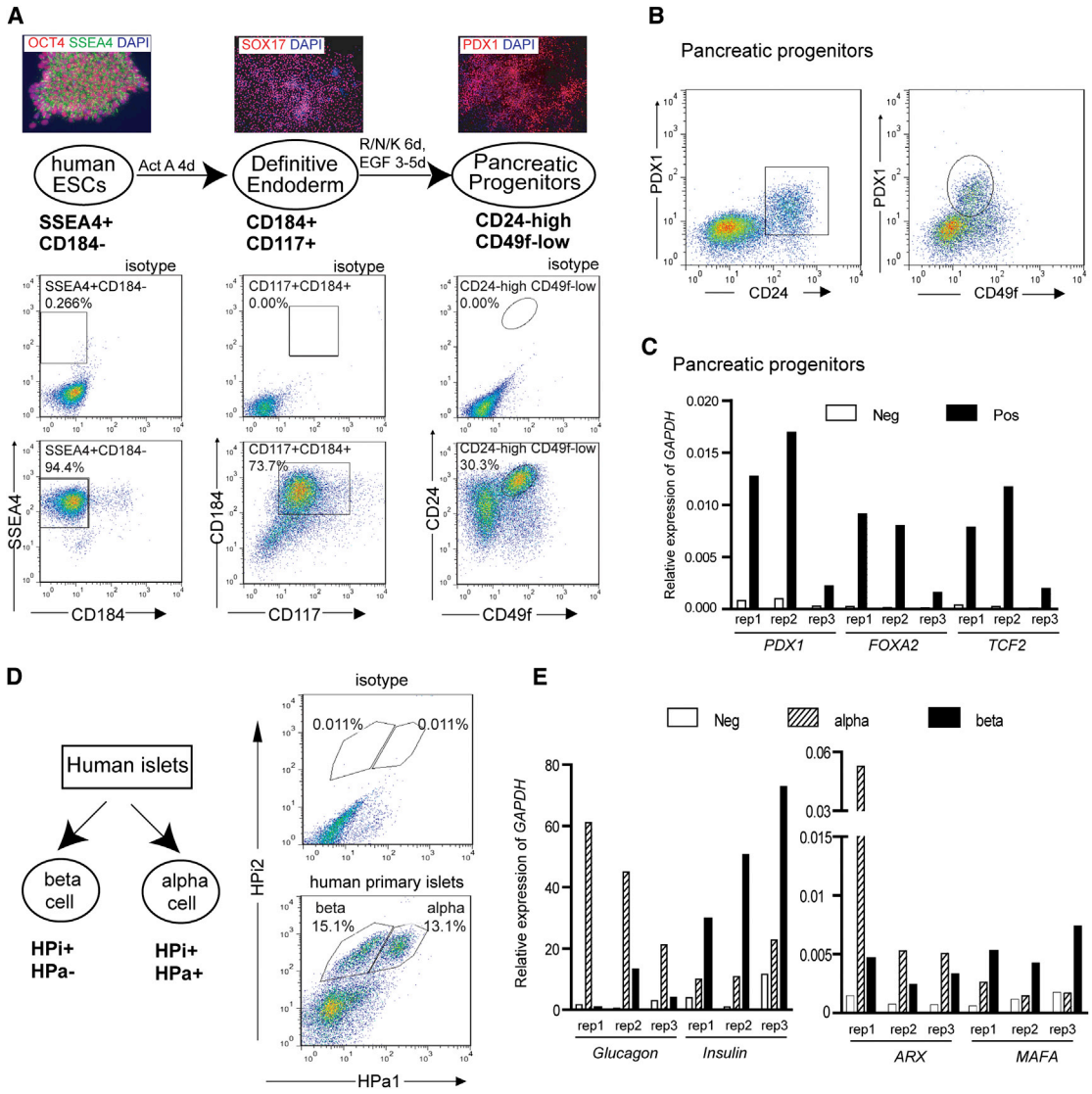


Figure 1. Purification of Samples for Transcriptome Analysis

(A) Purification of human ESCs, DE, and PPs. The top panel illustrates in vitro differentiation (R, retinoic acid; N, Noggin; K, KGF), immunostaining, and FACS sorting of human ESCs (SSEA4+/CD184−), DE (CD184+/CD117+), and PPs ($CD24^{high}$/$CD49f^{med}$).

(B) Flow-cytometric analysis demonstrates that PDX1-positive cell populations are $CD24^{high}$/$CD49f^{med}$.

(C) Comparison of the gene expression of key pancreatic transcriptional factors between $CD24^{high}$/$CD49f^{med}$ (Pos) and the remaining cells (Neg). Shown are three independent experiments.

(D) Purification of alpha and beta cells from human primary islets using HPi2 and HPa1. HPi2+/HPa1− mark beta cells, and HPi2+/HPa1+ mark alpha cells.

(E) qRT-PCR analysis of cell-type-specific marker gene expression confirms the identity of sorted alpha, beta, and non-alpha/beta cells. Shown are three independent experiments.

See also Figure S1 and Table S4.

2012) and Human Body Map intergenic lncRNA annotation (Cabili et al., 2011). Using stringent criteria (FPKM > 1 in each replicate), we identified 1,059 expressed intergenic lncRNAs that were not located within any annotated protein-coding genes (Table S1). Among these lncRNAs, 250 exhibit a stage-specific expression pattern (Figure 3A), 62 are ES specific, 82 are DE specific, 28 are PP specific, 55 are beta specific, and 23 are alpha specific. Representative stage-specific lncRNAs are shown in Figure 3C. Since lncRNAs likely co-express with their neighboring genes (Cabili et al., 2011), we attempted to obtain some insights into the functions of these lncRNAs by asking whether the neighboring protein-coding genes of these lncRNAs (genomic distance of <20 kb) are enriched for any biological processes. A Gene Ontology term analysis revealed that protein-coding genes close to the ES-specific lncRNAs are enriched for regulation of cell division and transcription, and genes close

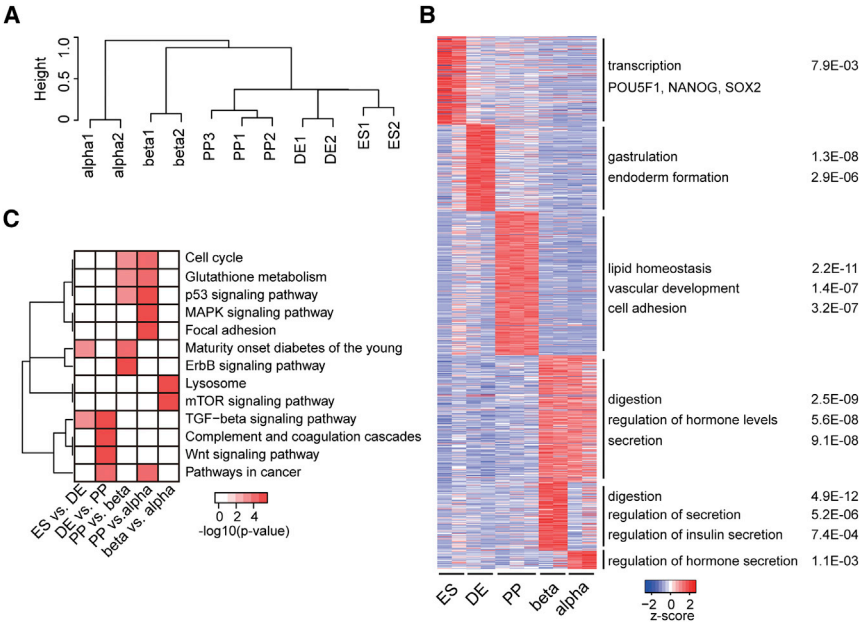


Figure 2. Expression Analysis of Protein-Coding Genes across Human Endoderm and Pancreatic Cell Lineages

(A) Cluster analysis of the transcriptome across human pancreatic cell lineages.

(B) Top enriched Gene Ontology terms of each group of stage-specific expressed genes are shown with gene counts and Fisher's exact test p value.

(C) Heatmap illustration of enriched KEGG signaling pathways based on pairwise comparisons of the various developmental stages (Fisher's exact test, p value < 0.01). Color scale represents the respective p value.

to DE-specific lncRNAs are enriched for endoderm development, dorsal/ventral pattern formation, and gastrulation (Figure 3B), suggesting a potential role for these lncRNAs in cell-fate determination.

Previous studies have shown that lncRNAs are highly tissue specific (Cabili et al., 2011; Morán et al., 2012). An analysis of the lncRNA expression profiles during endoderm and pancreatic lineage specification indicated that lncRNA expression is highly stage specific: more than 25% of lncRNAs in most of the stages exhibited a stage-specific expression pattern (13.9% for the PP stage), whereas less than 5% of the protein-coding genes were stage specific (Figure 3D). Consistent with previous reports (Cabili et al., 2011; Morán et al., 2012), statistically significant enrichment ($p < 1 \times 10^{-16}$ for each comparison, chi-square test) of lncRNAs over protein-coding genes was observed during human endoderm and pancreatic lineage specification. To investigate the potential role of these lncRNAs in differentiation, we examined the relationship between the expression of these lncRNAs and their neighboring protein-coding genes during endoderm and pancreatic lineage specification. The expression of lncRNAs and their neighboring genes was highly correlated in ESCs, DE, and PP (Figure 3E); however, the correlation between the expression of lncRNAs and their neighboring genes in terminally differentiated islet cells was less robust when compared with that of ES, DE, and PP samples (Figure 3E; $p = 8.3 \times 10^{-15}$ compared with ESCs, $p = 2.9 \times 10^{-16}$ compared with DE, $p = 4.2 \times 10^{-8}$ compared with PP cells; Wilcoxon rank sum test). These results suggest that lncRNAs in terminally differentiated cells are less likely to be co-regulated by their neighboring genes, indicating that they may serve specific functions in terminally differentiated islet cells.

DEANR1 Is a Human-DE-Specific lncRNA

lncRNAs have been reported to participate in many biological processes (Wang and Chang, 2011), but no lncRNA has been shown to be functionally important for human endoderm and pancreatic lineage specification. To begin to functionally characterize the hundreds of lncRNAs that exhibit stage-specific expression during endoderm and pancreatic lineage specification, we utilized a well-established DE differentiation system to test whether some of these lncRNAs have a role in DE differentiation. To this end, we applied more stringent criteria (higher FPKM > 5, log2 [fold change] > 1.5, p value < 0.05), and identified 38 lncRNAs that are differentially expressed in ESCs and DE cells (13 highly expressed in ESCs and 25 highly expressed in DE cells; Figure 4A; detailed information is provided in Table S2). Among the 25 DE-expressed lncRNAs, 12 are located physically close to protein-coding genes, including the endoderm factors GATA6 and FOXA2. Given that GATA6-associated lncRNAs are divergently transcribed, they likely share a promoter and thus may confound analysis. Therefore, we decided to focus on the FOXA2-associated lncRNA *LINC00261*. Because this is the first DE-specific lncRNA to be characterized, we renamed *LINC00261* as *DEANR1* (for definitive endoderm-associated long non-coding RNA1).

Our RNA-seq data indicated that *DEANR1* is highly expressed in DE, albeit to a lower level than FOXA2 (FPKM: 10 versus 35; Figure 4B). To further characterize the expression pattern of *DEANR1*, we analyzed its expression level in human ESCs, differentiated neuroectoderm (NE), differentiated embryoid body (EB), and DE cells by qRT-PCR. The results presented in Figure 4C indicate that the expression level of *DEANR1* in DE is at least 100-fold higher than in any of the other cell types analyzed. Furthermore, a time-course analysis demonstrated that the expression level of *DEANR1* during endoderm differentiation is gradually increased over differentiation (Figure 4D). Based on these results, we conclude that *DEANR1* is an lncRNA that is highly expressed in human DE.

DEANR1 Is Important for Endoderm Differentiation but Is Not Required for ESC Maintenance

To examine whether *DEANR1* is functionally important for endoderm differentiation, we generated two small hairpin RNA (shRNA) constructs targeting different regions of *DEANR1*. A qRT-PCR analysis demonstrated that both shRNAs achieved

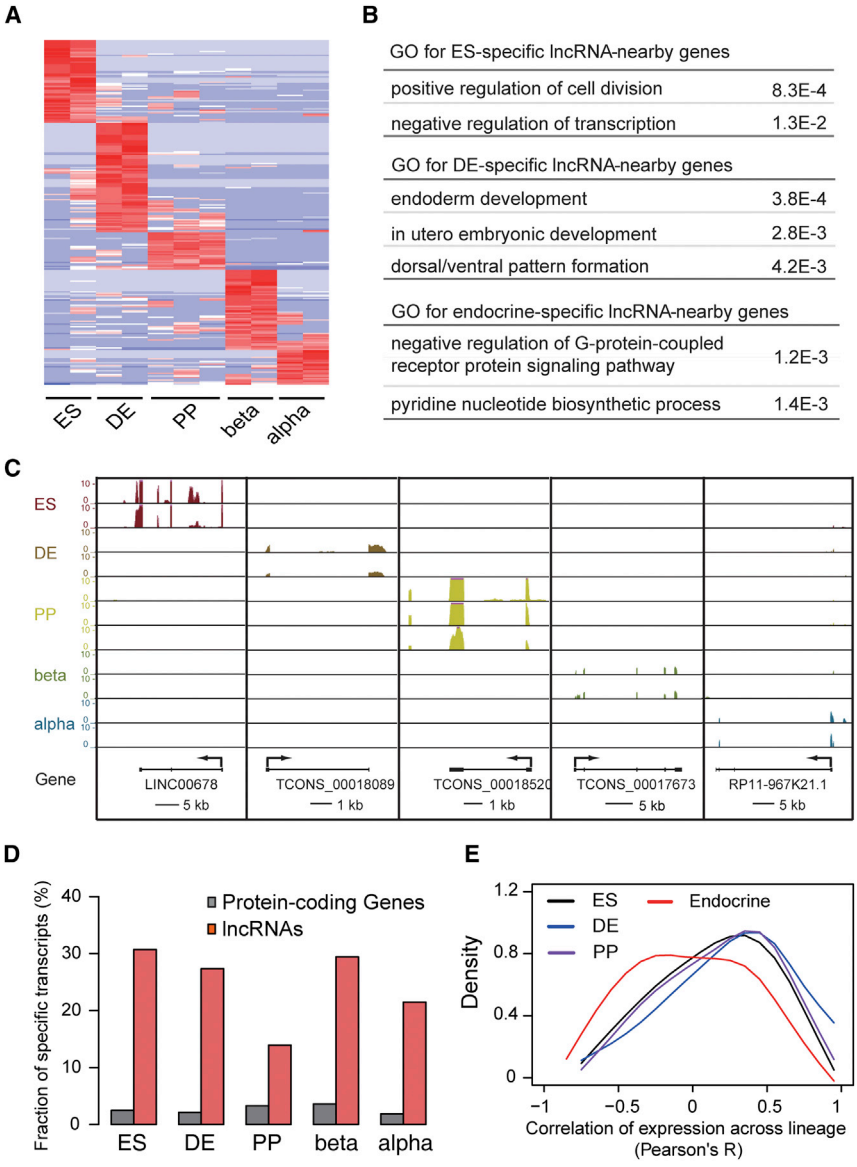


Figure 3. Dynamic Regulation of lncRNAs during Human Endoderm and Pancreatic Cell Lineage Specification

(A) Hierarchical clustering of 250 stage-specific lncRNAs during human pancreatic cell lineage specification. Red and blue represent high and low expression, respectively.

(B) Top enriched Gene Ontology terms for stage-specific lncRNA neighboring protein-coding genes.

(C) RNA-seq reads alignment of representative stage-specific expressed lncRNAs. For each lncRNA, the vertical axis is scaled at the same expression level for all samples. The lncRNA transcripts are depicted as black bars, with an arrow indicating the transcription direction.

(D) Relative percentage of stage-specific expressed lncRNAs (red bars) compared with protein-coding genes (gray bars) in five different human pancreatic cell lineage samples ($p < 1 \times 10^{-16}$ for each comparison, chi-square test).

(E) Endocrine islet-expressed lncRNAs show less correlation with the expression of their nearby genes compared with the lncRNAs in ES (black), DE (blue), and PP (purple) ($p = 8.3 \times 10^{-15}$ compared to ES, $p = 2.9 \times 10^{-16}$ compared with DE, $p = 4.2 \times 10^{-8}$ compared to PP cells; Wilcoxon rank sum test).

See also Table S1.

70%–75% knockdown of *DEANR1* in human ESCs (Figure S2A). Importantly, neither knockdown affected human ESC maintenance, as the expression level of pluripotent factors such as *SOX2*, *NANOG*, and *OCT4* was not altered in the stable knockdown cell lines (Figures S2B and S2C). In addition, neither the morphology nor the growth rate of the stable *DEANR1* knockdown cell lines was altered (data not shown). Thus, *DEANR1* does not seem to be required for human ESC maintenance, which is consistent with its low expression level in undifferentiated ESCs (Figures 4B and 4C).

To evaluate the effect of *DEANR1* knockdown on DE differentiation, we subjected stable *DEANR1* knockdown cell lines to endoderm differentiation conditions. A FACS analysis revealed that *DEANR1* knockdown resulted in a decrease in the percentage of CD184/CD117 double-positive cells compared with the control human ESCs (Figure 4E). Consistently, immunostaining results revealed that *DEANR1* knockdown cells exhibited a decreased expression of endoderm marker genes, such as *FOXA2* and *SOX17*, after being subjected to endoderm differentiation (Figure 4F).

To further analyze the role of *DEANR1* in endoderm differentiation, we performed RNA-seq of *DEANR1*-knockdown cells after subjecting them to endoderm differentiation (Table S3). Compared with differentiated control cells, 632 genes were downregulated and 569 genes were upregulated (Figure 4G) upon *DEANR1* knockdown (FDR < 0.001 and fold change > 2). These differentially expressed genes are significantly enriched for DE- and ES-signature genes. Among these, 219 downregulated genes in *DEANR1*-knockdown cells are DE-signature genes that are expected to be upregulated during endoderm differentiation (Figure 4G, top). Thus, the transcriptome analysis suggested a defective activation of endoderm genes in *DEANR1*-knockdown cells. A total of 194 upregulated genes in *DEANR1*-knockdown cells are ES-signature genes that should be downregulated during endoderm differentiation (Figure 4G, bottom). These transcriptome data, together with the differentiation data, suggest that *DEANR1* plays an important role in DE differentiation.

DEANR1 Regulates FOXA2 Expression

To identify potential targets of *DEANR1* in endoderm differentiation, we collected 47 RNA-seq data sets generated in this study and others (Bramswig et al., 2013; Morán et al.,

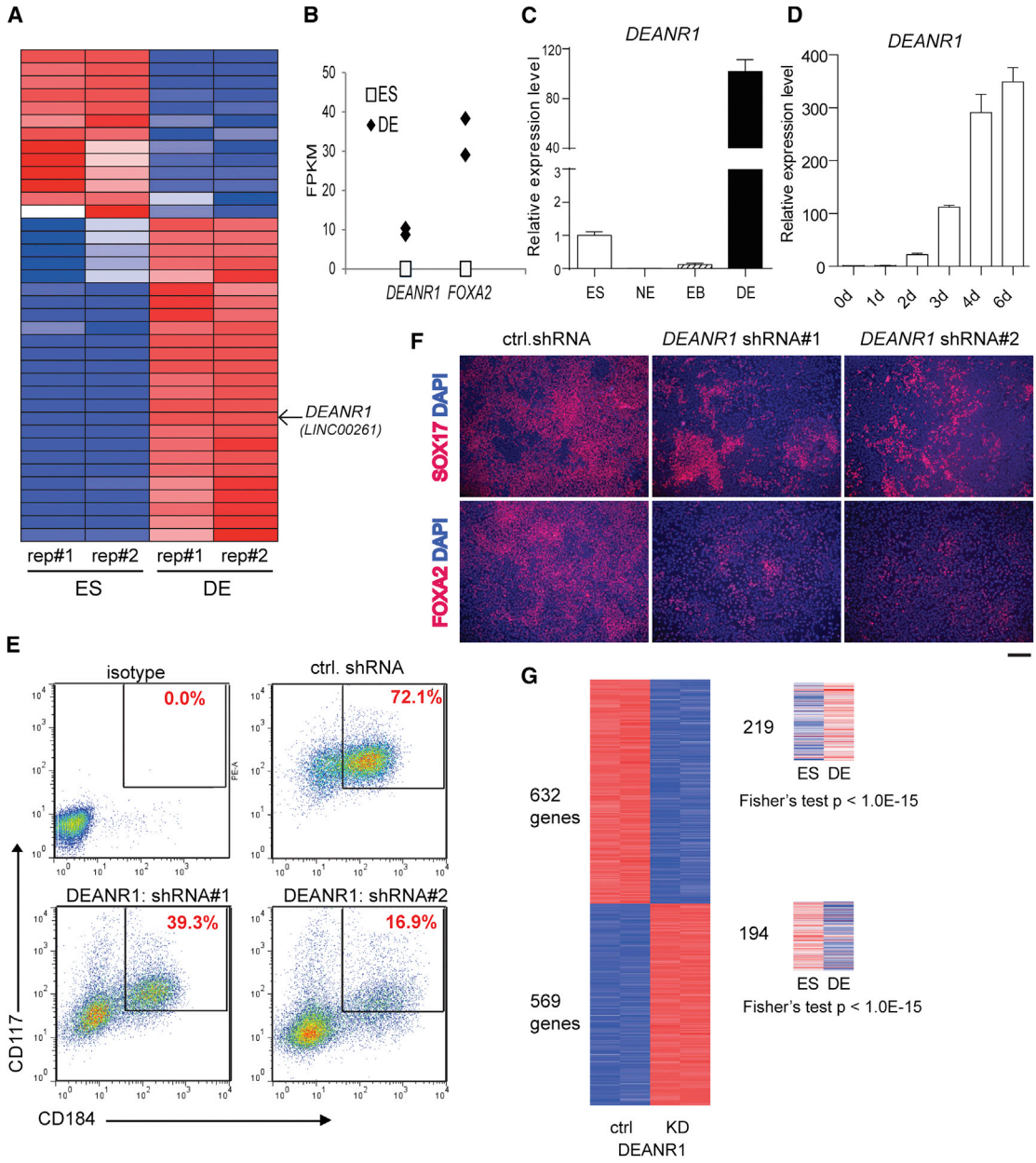


Figure 4. Identification of *DEANR1* as an Endoderm-Specific lncRNA

(A) Heatmap showing the expression levels of lncRNAs differentially expressed in ESCs and DE cells. Criteria: higher FPKM > 5, log2 (fold change) > 1.5, and p value < 0.05. Red and blue represent high and low expression, respectively. See also Table S2.

(B) Relative enrichment of *DEANR1* in DE compared with ESCs shown by FPKM. FOXA2 serves as a control.

(C) Relative expression level of *DEANR1* in various cell types quantified by qRT-PCR. Undifferentiated ESCs (ES), differentiated neuroectoderm (NE), spontaneously differentiated embryoid body (EB), and DE samples were compared. The expression level in ES was set as one.

(D) The expression dynamics of *DEANR1* during DE differentiation. The mean values are shown and error bars represent the SD from the mean (n = 3). The expression level at day 0 was set as one.

(E) Representative FACS analysis showing a reduced CD117/CD184 double-positive DE cell population upon *DEANR1* knockdown.

(F) Representative immunostaining of the DE markers SOX17 and FOXA2 in DE differentiated control and *DEANR1* knockdown cells. Scale bar represents 100 μm.

(G) RNA-seq analysis of differentiated *DEANR1*-knockdown endoderm cells and control endoderm cells. Compared with control (FDR < 0.001 and fold change > 2), 632 genes were downregulated (significantly enriched in DE-signature genes) and 569 genes were upregulated upon *DEANR1* knockdown (significantly enriched in ES-signature genes). Among these, 219 downregulated genes in *DEANR1*-knockdown cells are DE-signature genes that should be upregulated during endoderm differentiation (top), and 194 upregulated genes in *DEANR1*-knockdown cells are ES-signature genes that should be downregulated during endoderm differentiation (bottom).

See also Figure S2 and Tables S3 and S4.

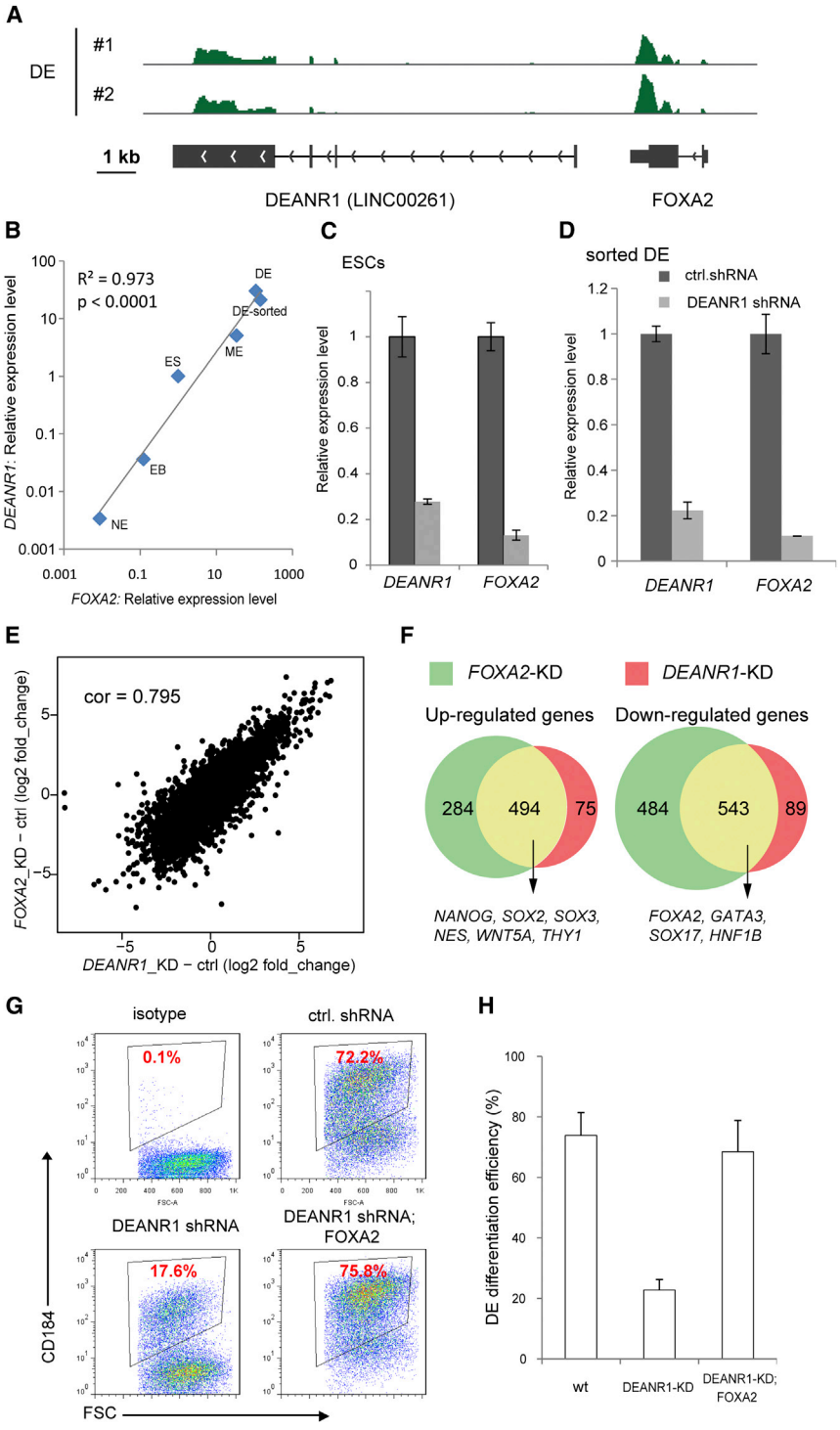


Figure 5. *FOXA2* Is a Major Functional Target of *DEANR1*

(A) Genomic location and diagrammatic presentation of RNA-seq results in DE for *DEANR1* and *FOXA2*.

(B) Correlation between *DEANR1* and *FOXA2* expression levels in human ESCs (ES), neuroectoderm (NE), embryoid body (EB), differentiated mesendoderm cells (ME), DE, and sorted DE cells.

(C) Knockdown of *DEANR1* in human ESCs reduces *FOXA2* expression. Relative qRT-PCR results are presented, with the *FOXA2* level in control knockdown ESCs set as one. shRNA#2 was used in (C)–(F).

(D) FACS-sorted *DEANR1*-knockdown DE cells exhibit decreased *FOXA2* expression.

(E) Scatterplot of the fold change of differentially expressed genes in *DEANR1*-knockdown and *FOXA2*-knockdown cells.

(F) Venn diagram showing that the genes affected by *DEANR1* knockdown largely overlap with those affected by FOXA2 knockdown. The common upregulated genes in both knockdowns include pluripotent genes (NANOG and SOX2) and ectoderm (SOX3 and NES) and mesoderm (WNT5A and THY1) marker genes. The common downregulated genes include endoderm-specific genes (FOXA2, GATA3, SOX17, and HNF1B).

(G) Representative FACS analysis results from control, *DEANR1*-knockdown, and *DEANR1*-knockdown rescue cells. The mean values are shown and error bars represent the SD from the mean (n = 3).

(H) Quantification of DE differentiation efficiency in control, *DEANR1*-knockdown, and *DEANR1*-knockdown rescue cells from three independent experiments.

See also Figure S3 and Table S4.

2012; Xie et al., 2013a, 2013b), and examined the correlation between the expression levels of *DEANR1* and transcription factors important for endoderm differentiation, such as *MIXL1*, *SOX17*, *FOXA2*, *GSC*, *EOMES*, and *GATA6*. We found that only expression of *FOXA2* was significantly correlated with that of *DEANR1* (Figure S3). Interestingly, *FOXA2* is also physically close to *DEANR1* in its genomic location (Figure 5A). Given that previous studies have demonstrated that lncRNAs can function in *cis* to regulate expression of adjacent protein-coding genes (Ørom et al., 2010), we examined the possibility that *DEANR1* may facilitate DE differentiation by regulating *FOXA2* expression. Consistent with this notion, we found that the expression level of *DEANR1* was highly correlated with the *FOXA2* level in different cell types (Figures 5B and S3). In addition, knockdown of *DEANR1* in human ESCs resulted in a dramatic decrease of *FOXA2* expression (Figure 5C). A similar effect was also observed in sorted CD184/CD117 double-positive endoderm cells derived from *DEANR1*-knockdown cells (Figure 5D). A gene-expression analysis in both undifferentiated ESCs (to exclude the differentiation effect) and purified endoderm cells (to bypass the endoderm differentiation defect) supported the notion that *DEANR1* regulates *FOXA2* expression.

To determine whether *FOXA2* is a major functional target of *DEANR1* in regulating DE differentiation, we first determined the transcriptome changes in *DEANR1*-knockdown cells and FOXA2-knockdown cells (80% knockdown efficiency; detailed information is provided in Table S3). Analysis of the RNA-seq data indicated that the changes induced by *DEANR1* knockdown and FOXA2 knockdown are highly correlated (Figure 5E). Moreover, the genes affected by *DEANR1* knockdown significantly overlap with those affected by FOXA2 knockdown (Figure 5F). The common upregulated genes in both knockdowns compared with control cells included pluripotent genes, ectoderm and mesoderm markers, and common downregulated genes included endoderm markers. These results suggest that either *DEANR1* knockdown or FOXA2 knockdown results in endoderm differentiation defects.

To further define the relationship between *DEANR1* and *FOXA2*, we attempted to rescue the DE differentiation defect of *DEANR1* knockdown by expressing exogenous *FOXA2.* To this end, we infected *DEANR1*-knockdown human ESCs with viruses expressing FOXA2, followed by DE differentiation. A FACS analysis demonstrated that the endoderm differentiation defect caused by *DEANR1* knockdown is successfully rescued by expression of *FOXA2* (Figures 5G and 5H), indicating that *FOXA2* is a major target that mediates the function of *DEANR1* in endoderm differentiation.

DEANR1 Contributes to *FOXA2* Transcription by Facilitating SMAD2/3 Recruitment

We next attempted to address how *DEANR1* contributes to *FOXA2* transcription. First, we examined the localization of the *DEANR1* RNA transcript. RNA fluorescent in situ hybridization (FISH) demonstrated that *DEANR1* is expressed and localized only to FOXA2-positive nuclei (Figure 6A). We noticed that *DEANR1* localized to a limited number of nuclear foci (two spots in most cases), which suggests that *DEANR1* has limited targets. We also checked the genomic location of the *DEANR1* transcript by RNA-DNA dual FISH, and the results showed that the *DEANR1* transcript is located in the FOXA2 gene locus (Figure 6B). Taken together, these data suggest that *DEANR1* likely contributes to *FOXA2* transcription in *cis*.

Previous studies demonstrated that *FOXA2* expression could be activated by several transcriptional factors, including SOX17 (Sinner et al., 2004), EOMES (Teo et al., 2011), and SMAD2/3 (Kim et al., 2011), during endoderm differentiation. Therefore, we asked whether any of these transcription factors can associate with *DEANR1*. Immunoprecipitation followed by qRT-PCR analysis demonstrated that only the SMAD2/3 antibody significantly enriched *DEANR1* (Figure 6C). The enrichment of *DEANR1* by the SMAD2/3 antibody is specific, as the antibody did not enrich for either the U1 small nuclear RNA (snRNA) or another lncRNA, *MALAT1* (Figure 6C). These data suggest that *DEANR1* may contribute to *FOXA2* activation by associating with SMAD2/3.

To examine whether binding of SMAD2/3 to the *FOXA2* promoter requires *DEANR1*, we first mapped the location of SMAD2/3 on the *FOXA2* gene locus. Chromatin immunoprecipitation (ChIP) assays demonstrated that SMAD2/3 binds to the promoter of *FOXA2*, but is absent from the upstream proximal region (Figure 6D). Importantly, knockdown of *DEANR1* reduced the binding of SMAD2/3 to the *FOXA2* promoter region, but had no significant effect on the binding of SMAD2/3 to another SMAD2/3 target, *EOMES* (Figure 6E). These results suggest that *DEANR1* associates with SMAD2/3 and facilitates its recruitment to the *FOXA2* promoter to activate *FOXA2* expression.

DISCUSSION

In this study, we performed a comprehensive transcriptome analysis of human endoderm and pancreatic lineage using highly purified samples (Figure 1). We identified hundreds of stage-specific lncRNAs and characterized the function of *DEANR1* in endoderm differentiation. Our data suggest that *DEANR1* plays an important role in regulating endoderm differentiation. The fact that overexpression of *FOXA2* can largely rescue the endoderm differentiation defect caused by *DEANR1* depletion suggests that *FOXA2* is a major functional target of *DEANR1*. Furthermore, our data support the notion that *DEANR1* regulates FOXA2 transcription in *cis*. *DEANR1* associates with SMAD2/3 and facilitates its binding to the *FOXA2* gene promoter to activate *FOXA2* expression (Figure 6F), which in turn promotes endoderm differentiation.

Our comprehensive transcriptome analysis of pancreatic beta-cell lineage specification provides insight into the transcriptional dynamics underlying cell-fate conversion. First, we observed a decrease in the number of actively transcribed genes during differentiation, particularly in the terminally differentiated islet cell types (Figure S1C), similar to what was observed in a previous study regarding neural differentiation (Wu et al., 2010). Together, the two studies suggest that restriction in gene expression might be a common phenomenon during cell-fate specialization. How to silence undesired gene expression during differentiation should be an important question for future investigations. Indeed, two recent studies that profiled the transcriptome and histone K4me3 and K27me3 using human ESC-derived pancreatic intermediate (Xie et al., 2013a) and primary islet cell types (Bramswig et al., 2013) revealed that silencing of transitory genes during lineage progression is associated with reinstatement of Polycomb-dependent repression. Furthermore, DNA methylation-mediated transcription repression has been reported to be essential for maintaining pancreatic beta cell identity (Dhawan et al., 2011). Thus, it is likely that both DNA methylation and H3K27 methylation play an important role in silencing undesired gene expression to achieve cell lineage restriction.

lncRNAs have been reported to participate in diverse biological processes, including ESC maintenance, neural cell differentiation, muscle differentiation, and hematopoiesis (Hu et al., 2012), as well as cardiac development (Grote et al., 2013; Klattenhoff et al., 2013) and epidermal differentiation (Kretz et al., 2013). Although a recent study reported the expression of thousands of lncRNAs during endodermal differentiation (Sigova et al., 2013), no specific function has been assigned to any of these lncRNAs during endoderm differentiation. Future characterization of the other endoderm-specific and pancreatic-lineage-specific lncRNAs identified in our study may reveal additional lncRNAs that are important for endoderm and pancreatic

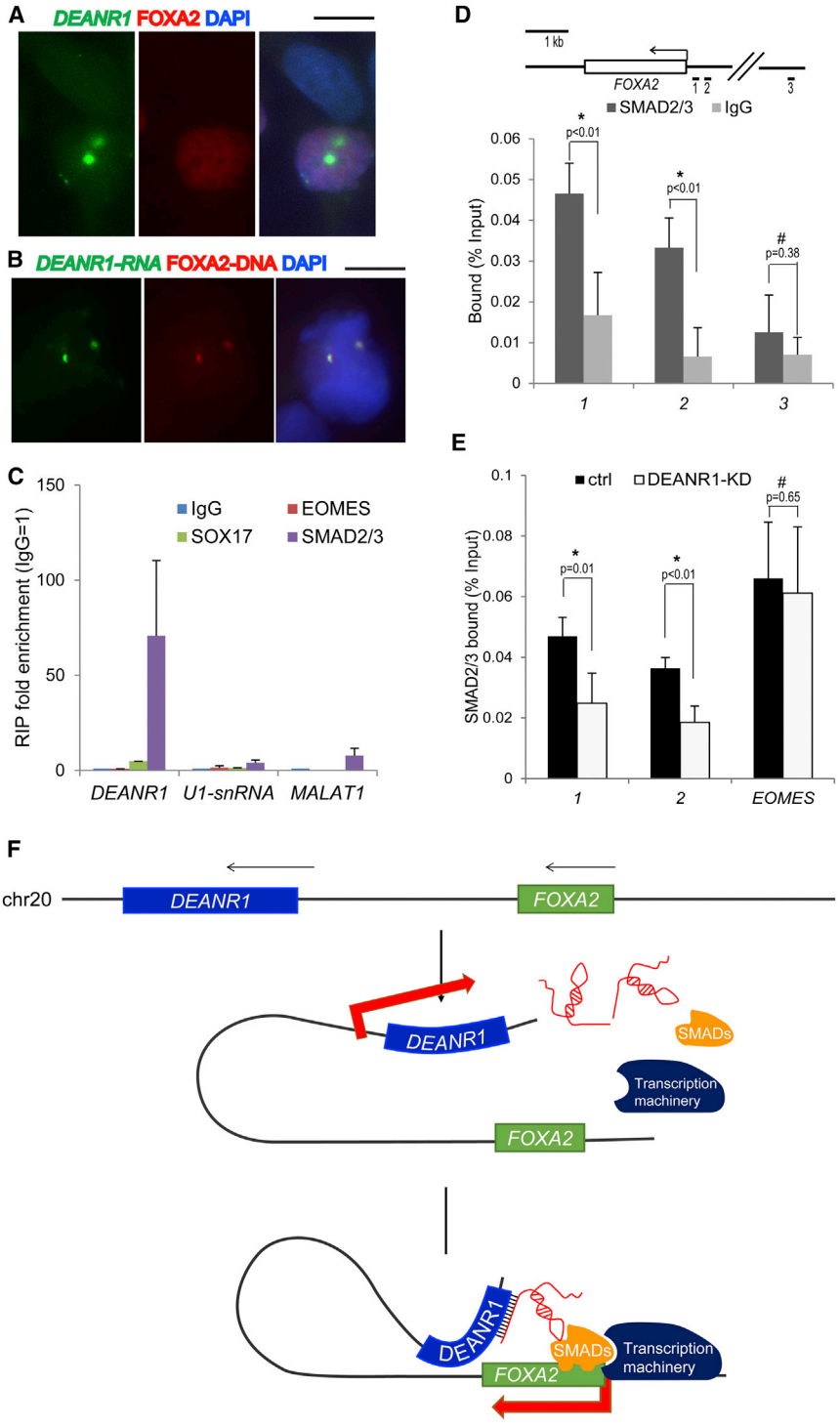


Figure 6. *DEANR1* Associates with SMAD2/3 and Helps to Target It to the *FOXA2* Promoter

(A) RNA-FISH demonstrates nuclear localization of *DEANR1* only in FOXA2-positive DE cells. Scale bar represents 5 μm.

(B) Dual RNA-DNA-FISH demonstrates that *DEANR1* transcripts (green signal) are localized to the *FOXA2* gene locus (red signal).

(C) SMAD2/3 antibodies immunoprecipitate *DEANR1*, but not U1 snRNA or lncRNA *MALAT1*.

(D) ChIP analysis demonstrates binding of SMAD2/3 to the promoter region of the *FOXA2* gene locus. IgG serves as a control. The genomic location of the analyzed regions is indicated in the diagram at the top of the panel.

(E) Knockdown of *DEANR1* reduces binding of SMAD2/3 to the *FOXA2* promoter, but not to another SMAD2/3 target, *EOMES*. The mean values are shown and error bars represent the SD from the mean (n = 3). Significance was determined by two-tailed Student's t test, and the p value is presented.

(F) Hypothetical model illustrating how *DEANR1* might regulate *FOXA2* transcription in *cis*. The genomic locations of *DEANR1* and *FOXA2* genes are shown on the top line and the arrows indicate the transcription direction. Upon *DEANR1* activation, chromatin looping brings the *DEANR1* locus spatially close to the *FOXA2* promoter region (~20 kb between the *DEANR1* gene body and the *FOXA2* promoter). The transcribed *DEANR1* interacts with SMAD2/3 protein and brings SMAD2/3 to the *FOXA2* promoter by forming DNA-RNA hydride for specific targeting of SMAD2/3 to the FOXA2 promoter. Together with other transcription machinery, binding of SMAD2/3 to the *FOXA2* promoter initiates transcription.

See also Table S4.

lineage specification. In addition, lncRNAs have been shown to act both in *cis* and in *trans* (Rinn and Chang, 2012; Ulitsky and Bartel, 2013). The observation that more than 60% of endoderm-expressed lncRNAs exhibit coordinated expression changes with their adjacent protein-coding genes (Sigova et al., 2013) suggests that *cis* regulation or local interaction between lncRNAs and their neighboring protein-coding genes may be a common feature of endoderm differentiation. The demonstration that *DEANR1* contributes to DE differentiation by regulating its neighboring gene, *FOXA2*, is consistent with this notion.

Nodal/activin signaling activates the SMAD2/3 pathway and promotes endoderm differentiation (Zorn and Wells, 2009). Secreted Nodal binds and activates ALK4/7, which in turn activates cytosolic SMAD2 and/or SMAD3 by phosphorylation. Phosphorylated SMAD2/3 then enters the nucleus to activate its downstream targets. While the mechanism of SMAD2/3 activation is conserved and well documented, the mechanisms by which SMAD2/3 activate their downstream target genes appears to be cell-context dependent (Feng and Derynck, 2005; Massagué et al., 2005). Although the MH1 domain of SMAD3 is capable of binding to DNA through Smad-binding elements

with weak affinity, native SMAD target promoters seldom contain such Smad-binding elements, and thus many partners, including DNA-binding proteins, have been reported to form complexes with SMAD2/3 to activate SMAD2/3 target gene expression (Feng and Derynck, 2005; Massagué et al., 2005). In our study, we demonstrated that the association of *DEANR1* with SMAD2/3 can facilitate SMAD2/3 binding to the *FOXA2* promoter. Based on the data we have obtained so far, we propose a hypothetical model that may explain how *DEANR1* contributes to FOXA2 activation and endoderm differentiation (Figure 6F). It is possible that activation of *DEANR1* by an endoderm differentiation signal causes chromatin looping to make the *DEANR1* locus spatially close to the FOXA2 promoter. The *DEANR1* transcripts can help bring SMAD2/3 to the FOXA2 promoter through a specific interaction with SMAD2/3 and at the same time base pair with a certain region of DNA that encodes *DEANR1*. Together with the recruitment of other transcription machinery, transcription of FOXA2 can be initiated. We cannot rule out the possibility that other proteins are also required for the binding of *DEANR1* to SMAD2/3, since no report has suggested that Smad2 is associated with an intrinsic RNA-binding property, and some RNA-binding proteins have been shown to be able to physically interact with Smad2, Smad3, and Smad4 (Sun et al., 2006). However, adenomatous polyposis coli (APC), a well-studied protein that is involved in the microtubule and WNT signaling pathway, was recently reported to be an RNA-binding protein (Preitner et al., 2014). Nevertheless, RNA-mediated SMAD2/3 targeting represents an alternative mechanism by which SMAD2/3 are recruited to their targets.

EXPERIMENTAL PROCEDURES

Sample Preparation for RNA-Seq

The following antibodies (all purchased from BD Biosciences) were used to label different cell populations for flow cytometry on a FACSAria II with FACSDiva software (version 6.0; BD Biosciences) as described previously (Dorrell et al., 2011): SSEA4+/CD184− for undifferentiated human ESCs; CD184+/CD117+ for DE; $CD24^{high}/CD49f^{med}$ for PPs; HPi2+/HPa1− for beta cells; and HPi2+/HPa1+ for alpha cells. DE cells and PPs were differentiated as previously described (Jiang et al., 2011) with minor modifications. Two individual islet preparations were used to generate alpha1/beta1 and alpha2/beta2 samples for sequencing.

RNA Deep Sequencing and Data Analysis

Total RNA was purified from 1,000 sorted cells using a ZR RNA microprep kit (Zymo Research). cDNA synthesis and amplification were performed with the SMARTer ultra-low input RNA kit (Clontech). Amplified cDNA (10–40 ng) was then fragmented by a Covaris S2 sonicator (Covaris) and converted to sequencing libraries with barcode in an Illumina Hiseq 2000 instrument (pair-end 100 bp). All obtained reads from each sample were mapped against the human genome (hg19 build) with Bowtie/Tophat v2.0.2, which allows mapping across splice sites by reads segmentation (Trapnell et al., 2012). The uniquely mapped reads were subsequently assembled into transcripts guided by reference annotation (Gencode v14 and RefSeq gene models) (Harrow et al., 2012; Pruitt et al., 2012) with Cufflinks v2.0.2 (Trapnell et al., 2012). The expression level of each gene was quantified with normalized FPKM (fragments per kilobase of exon per million mapped fragments). Functional annotation of significantly different transcripts and enrichment analysis were performed with DAVID (Huang et al., 2009).

To quantify the stage specificity of genes and lncRNAs, the expression levels of the samples in a certain stage versus the other samples in human endoderm and pancreatic lineage were compared. The islet-specificity score of each transcript was defined as follows:

$$score = meanA - (meanOther + 2 * sdOther),$$

where meanA is the mean expression of the samples in certain stage, and meanOther and sdOther are the mean and SD of the expression levels in the other samples, respectively. Therefore, a positive score indicated that the gene or lncRNAs was expressed in a certain stage at a considerably higher level than in the rest of the stages. A gene or lncRNA with a score of >0.5 was considered as specifically expressed in a certain stage.

Genetic Manipulation

The human ESC line HUES8, cultured without feeder, was used to generate stable knockdown cell lines. A lentiviral vector system containing the murine U6-shRNA cassette and a puromycin selection marker was used to achieve high transfection efficiency. We used two efficient shRNA targeting sites against *DEANR1* and one against *FOXA2*, as follows: DEANR1-shRNA1, GAAAGCTGTAGCCATTCAA (position: 761–779 nt); DEANR1-shRNA2, GCAATTAATTCAGGACACT (position: 2,979–2,997 nt); and FOXA2-shRNA, GCAAGGGAGAAGAAATCCATA.

FISH

Sequential protein staining and RNA detection were performed as previously described (Namekawa and Lee, 2011; Takizawa et al., 2008). In brief, RNA-FISH was performed by using a nick-translated probe (Roche) followed by a tyramide signal amplification kit (Life Technologies). After RNA-FISH, the cells were treated by RNase A and denatured. Nick-translated BAC containing FOXA2 (RP11-842A16) was labeled as a DNA probe.

RNA-Binding Protein Immunoprecipitation and ChIP

An RNA-binding protein immunoprecipitation (RIP) assay was performed using the Magna RIP kit (Millipore) according to the manufacturer's instruction. Cell lysates from 10 million cells and 2–5 μg of control IgG or antibody against SOX17 (R&D, AF1924), EOMES (Abcam, ab23345), or SMAD2/3 (R&D, AF3797) were used. We validated the RIP assay using the SNRNP70 antibody, which can bind to U1 snRNA. A ChIP assay was performed using magnetic beads. Equal amounts of chromatin from each sample and 1 μg of control IgG or antibody against SMAD2/3 were used.

ACCESSION NUMBERS

The data reported in this paper have been deposited to the NCBI GEO and are available under accession number GSE44875.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.03.008.

AUTHOR CONTRIBUTIONS

W.J. and Y.Z. conceived the project, designed the experiments, and wrote the manuscript. W.J. performed the experiments. Y.L. performed the bioinformatics analysis. R.L. prepared the low-input library and helped with the bioinformatics analysis. K.Z. assisted in analysis and data interpretation.

ACKNOWLEDGMENTS

We thank Dr. Philip Streeter (Oregon Stem Cell Center) for providing the antibodies against HPi2 and HPa1. We also thank Drs. Donghui Zhang and Shinpei Yamaguchi for their help in making Figure 6F, Drs. Azusa Inoue and Shogo Matoba for advice on RNA-FISH, and Dr. Luis M. Tuesta for a critical reading of the manuscript. This work was supported by NIH grant U01-DK089565 (Y.Z.). W.J. is supported by a Juvenile Diabetes Research Foundation postdoctoral

fellowship (3-2012-293). Y.Z. is an Investigator of the Howard Hughes Medical Institute.

Received: January 21, 2015
Revised: February 12, 2015
Accepted: February 27, 2015
Published: April 2, 2015

REFERENCES

Bramswig, N.C., Everett, L.J., Schug, J., Dorrell, C., Liu, C., Luo, Y., Streeter, P.R., Naji, A., Grompe, M., and Kaestner, K.H. (2013). Epigenomic plasticity enables human pancreatic α to β cell reprogramming. J. Clin. Invest. *123*, 1275–1284.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al.; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559–1563.

Cheng, X., Ying, L., Lu, L., Galvão, A.M., Mills, J.A., Lin, H.C., Kotton, D.N., Shen, S.S., Nostro, M.C., Choi, J.K., et al. (2012). Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. Cell Stem Cell *10*, 371–384.

D'Amour, K.A., Bang, A.G., Eliazer, S., Kelly, O.G., Agulnick, A.D., Smart, N.G., Moorman, M.A., Kroon, E., Carpenter, M.K., and Baetge, E.E. (2006). Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. Nat. Biotechnol. *24*, 1392–1401.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

Dhawan, S., Georgia, S., Tschen, S.I., Fan, G., and Bhushan, A. (2011). Pancreatic β cell identity is maintained by DNA methylation-mediated repression of Arx. Dev. Cell *20*, 419–429.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Dorrell, C., Abraham, S.L., Lanxon-Cookson, K.M., Canaday, P.S., Streeter, P.R., and Grompe, M. (2008). Isolation of major pancreatic cell types and long-term culture-initiating cells using novel human surface markers. Stem Cell Res. (Amst.) *1*, 183–194.

Dorrell, C., Schug, J., Lin, C.F., Canaday, P.S., Fox, A.J., Smirnova, O., Bonnah, R., Streeter, P.R., Stoeckert, C.J., Jr., Kaestner, K.H., and Grompe, M. (2011). Transcriptomes of the major human pancreatic cell types. Diabetologia *54*, 2832–2844.

Feng, X.H., and Derynck, R. (2005). Specificity and versatility in tgf-beta signaling through Smads. Annu. Rev. Cell Dev. Biol. *21*, 659–693.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., and Herrmann, B.G. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev. Cell *24*, 206–214.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Hu, W., Alvarez-Dominguez, J.R., and Lodish, H.F. (2012). Regulation of mammalian cell differentiation by long non-coding RNAs. EMBO Rep. *13*, 971–983.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Jiang, W., Shi, Y., Zhao, D., Chen, S., Yong, J., Zhang, J., Qing, T., Sun, X., Zhang, P., Ding, M., et al. (2007). In vitro derivation of functional insulin-producing cells from human embryonic stem cells. Cell Res. *17*, 333–344.

Jiang, W., Sui, X., Zhang, D., Liu, M., Ding, M., Shi, Y., and Deng, H. (2011). CD24: a novel surface marker for PDX1-positive pancreatic progenitors derived from human embryonic stem cells. Stem Cells *29*, 609–617.

Jiang, W., Wang, J., and Zhang, Y. (2013a). Histone H3K27me3 demethylases KDM6A and KDM6B modulate definitive endoderm differentiation from human ESCs by regulating WNT signaling pathway. Cell Res. *23*, 122–130.

Jiang, W., Zhang, D., Bursac, N., and Zhang, Y. (2013b). WNT3 is a biomarker capable of predicting the definitive endoderm differentiation potential of hESCs. Stem Cell Reports *1*, 46–52.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al.; RIKEN Genome Exploration Research Group; Genome Science Group (Genome Network Project Core Group); FANTOM Consortium (2005). Antisense transcription in the mammalian transcriptome. Science *309*, 1564–1566.

Kim, S.W., Yoon, S.J., Chuong, E., Oyolu, C., Wills, A.E., Gupta, R., and Baker, J. (2011). Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. Dev. Biol. *357*, 492–504.

Klattenhoff, C.A., Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P.A., Steinhauser, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S., et al. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell *152*, 570–583.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature *493*, 231–235.

Kroon, E., Martinson, L.A., Kadoya, K., Bang, A.G., Kelly, O.G., Eliazer, S., Young, H., Richardson, M., Smart, N.G., Cunningham, J., et al. (2008). Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nat. Biotechnol. *26*, 443–452.

Massagué, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev. *19*, 2783–2810.

McKnight, K.D., Wang, P., and Kim, S.K. (2010). Deconstructing pancreas development to reconstruct human islets from pluripotent stem cells. Cell Stem Cell *6*, 300–308.

Morán, I., Akerman, I., van de Bunt, M., Xie, R., Benazra, M., Nammo, T., Arnes, L., Nakić, N., García-Hurtado, J., Rodríguez-Seguí, S., et al. (2012). Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. Cell Metab. *16*, 435–448.

Namekawa, S.H., and Lee, J.T. (2011). Detection of nascent RNA, single-copy DNA and protein localization by immunoFISH in mouse germ cells and preimplantation embryos. Nat. Protoc. *6*, 270–284.

Nostro, M.C., Sarangi, F., Ogawa, S., Holtzinger, A., Corneo, B., Li, X., Micallef, S.J., Park, I.H., Basford, C., Wheeler, M.B., et al. (2011). Stage-specific signaling through TGFβ family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. Development *138*, 861–871.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46–58.

Pagliuca, F.W., Millman, J.R., Gürtler, M., Segel, M., Van Dervort, A., Ryu, J.H., Peterson, Q.P., Greiner, D., and Melton, D.A. (2014). Generation of functional human pancreatic β cells in vitro. Cell *159*, 428–439.

Paige, S.L., Thomas, S., Stoick-Cooper, C.L., Wang, H., Maves, L., Sandstrom, R., Pabon, L., Reinecke, H., Pratt, G., Keller, G., et al. (2012). A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell *151*, 221–232.

Preitner, N., Quan, J., Nowakowski, D.W., Hancock, M.L., Shi, J., Tcherkezian, J., Young-Pearse, T.L., and Flanagan, J.G. (2014). APC is an RNA-binding protein, and its interactome provides a link to neural development and microtubule assembly. Cell *158*, 368–382.

Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. *40*, D130–D135.

Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat. Biotechnol. *32*, 1121–1133.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M., et al. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. eLife *2*, e01749.

Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C., and Young, R.A. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc. Natl. Acad. Sci. USA *110*, 2876–2881.

Sinner, D., Rankin, S., Lee, M., and Zorn, A.M. (2004). Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. Development *131*, 3069–3080.

Sun, Y., Ding, L., Zhang, H., Han, J., Yang, X., Yan, J., Zhu, Y., Li, J., Song, H., and Ye, Q. (2006). Potentiation of Smad-mediated transcriptional activation by the RNA-binding protein RBPMS. Nucleic Acids Res. *34*, 6314–6326.

Takizawa, T., Gudla, P.R., Guo, L., Lockett, S., and Misteli, T. (2008). Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. Genes Dev. *22*, 489–498.

Teo, A.K., Arnold, S.J., Trotter, M.W., Brown, S., Ang, L.T., Chng, Z., Robertson, E.J., Dunn, N.R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes Dev. *25*, 238–250.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Mol. Cell *43*, 904–914.

Wang, P., Xue, Y., Han, Y., Lin, L., Wu, C., Xu, S., Jiang, Z., Xu, J., Liu, Q., and Cao, X. (2014). The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. Science *344*, 310–313.

Wu, J.Q., Habegger, L., Noisa, P., Szekely, A., Qiu, C., Hutchison, S., Raha, D., Egholm, M., Lin, H., Weissman, S., et al. (2010). Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. Proc. Natl. Acad. Sci. USA *107*, 5254–5259.

Xie, R., Everett, L.J., Lim, H.W., Patel, N.A., Schug, J., Kroon, E., Kelly, O.G., Wang, A., D'Amour, K.A., Robins, A.J., et al. (2013a). Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. Cell Stem Cell *12*, 224–237.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013b). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134–1148.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. Nature *492*, 443–447.

Zhang, D., Jiang, W., Liu, M., Sui, X., Yin, X., Chen, S., Shi, Y., and Deng, H. (2009). Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells. Cell Res. *19*, 429–438.

Zorn, A.M., and Wells, J.M. (2009). Vertebrate endoderm development and organ formation. Annu. Rev. Cell Dev. Biol. *25*, 221–251.